

LESSON PLAN

Unpacking the Black Box: Explaining Algorithms and AI

This lesson is part of *USE, UNDERSTAND & ENGAGE: A Digital Media Literacy Framework for Canadian Schools*: <http://mediasmarts.ca/teacher-resources/digital-literacy-framework>.



LEVEL: Grade 9 to 12

DURATION: 3 1/2 – 4 hours

ABOUT THE AUTHOR: Melissa Racine, Media Education Specialist, MediaSmarts

Overview

In this lesson, students learn about algorithms and AI, how they work, how they impact our lives on the internet, and ethical considerations. The lesson begins with a class discussion on algorithms. Students will discuss how AIs reinforce real-world biases, the difficulties in identifying how AIs make decisions, what information algorithms use to make choices, and how that information impacts the types of decisions AIs make. Finally, students will demonstrate their knowledge by researching and designing an infographic on a field that uses algorithms to make decisions. This lesson aims to build critical thinking skills by examining how AI algorithms work, investigating the biases and impacts of AI decision-making, and reflecting on how the implications to their own lives.

Learning Outcomes

Big ideas/key concepts: Students will learn understand that...

- Media have social and political implications:
 - Algorithms are used to make important decisions
 - Algorithms reproduce and can intensify existing social biases and stereotypes
- Media have commercial considerations:
 - Companies optimize algorithms to keep users engaged
- Digital media have unanticipated audiences:
 - Data collected about us influences algorithmic decisions
- Digital media experiences are shaped by the tools we use:
 - How we use digital tools is influenced by algorithm design
- Interactions through digital media have real impact:
 - How we use platforms is influenced by algorithmic design

Key questions:

- How do algorithms work?
- How does generative artificial intelligence work?
- How do algorithms and artificial intelligence affect our lives?
- What can we do about it?

Frequent misconceptions to correct:

- Algorithms are “just math” and cannot be biased
- We are only affected by data collection if platforms learn something specific about us

Essential knowledge: Students will learn...

- Reading media: How algorithms and generative AI work
- Consumer awareness: How platforms use algorithms to keep users engaged
- Community engagement: How algorithms can be biased and lead to unfair or unexplained decisions
- Privacy and security: How data collected about us, and inferred based on collected data, can influence algorithmic decisions

Performance tasks: Students will...

- Use: Create an informational media work
- Understand: Analyze how algorithms and generative AI work
- Engage: Evaluate the impacts of algorithms and generative AI on themselves and society

Preparation and Materials

Review and prepare to project the [Unpacking the Black Box: Explaining Algorithms and AI](#) slideshow

Prepare to distribute the *Algorithm AIs and Bias Assignment Sheet*

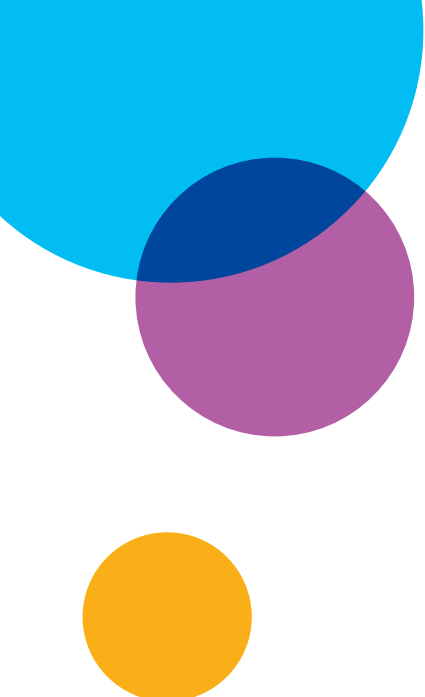
Prepare to distribute the *Algorithm AIs and Bias Planning Sheet*

Optional: Review the MediaSmarts article on [Fair Dealing for Media Education](#)

Procedure

WHAT DO YOU KNOW ABOUT ALGORITHMS?

Show **slide 1** of the *Unpacking the Black Box* slideshow and ask students what they think they know about algorithms. Give some time for them to



offer a few ideas, then advance to **slide 2** and ask them how they think algorithms work and how they impact people's lives.

Show **slide 3** and tell students that you're going to play a game. The letter on the slide is the first letter of a word. Allow students to guess what the word is and ask how confident they are in that guess.

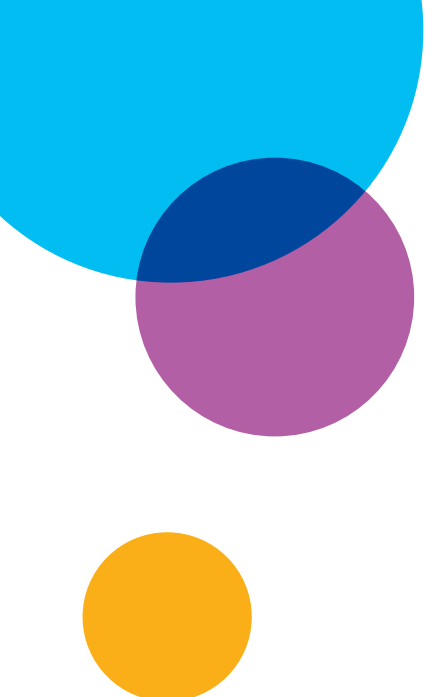
Advance through **slides 4-6**. As each new letter is added, ask students to note how their guesses – and their confidence – have changed. As the word gets longer, students are likely narrowing the answer down to just a few possibilities.

Advance to **slide 7**. Ask students if they guessed what the word was before getting to this slide. For those who did, ask them to explain their thought process as to how they guessed the word. Follow this up by asking what extra information could have helped them guess the word sooner (some examples include if they knew that you liked hippos or knowing that it was going to be the name of an animal). Tell students that this might remind them of an algorithm that they probably use every day: the autocorrect function on their phone. Explain that autocorrect uses patterns to guess what word they are trying to type, and its guesses are more accurate with the more letters they type and if they're typing a common word.

PURPOSES OF ALGORITHMS

Show **slide 8** and explain that algorithms are a series of steps or instructions for doing something. Algorithms are used for *sorting* things, like search results; *classifying* things into different groups; *matching* different things and *filtering* out before you see them.

Advance to **slide 9**. Connect this definition to a relevant, real-world example – social media. For those with social network accounts, the company has already sorted them based on what they know, or think they know, about their age, gender, interests, and dozens or even hundreds of other bits of data. They use this to classify their users into an “audience” – a group of people with similar tastes or interests – and to match them with recommended posts or videos and targeted ads. They also use the same data to filter out content that their data calculates they won't be interested in, ads they probably won't respond to, and content that's against the terms of service.



Progress to **slide 10**. Point out that this may not seem so bad. After all, if you have to see ads, it's better to see ones for things you're actually interested in. But because algorithms show us what they think we want to see, they can keep us from seeing the whole picture. You may miss an important post from one of your friends because the algorithm doesn't think you'll like it. You may not get the best or most reliable results from a search engine because its algorithm thinks you'll be more interested in different sources. Algorithms on social networks and video sites also usually prefer whatever people have interacted with the most, which means that hoaxes, conspiracy theories and misinformation often spread more easily than reliable information and the loudest voices can seem like the majority.

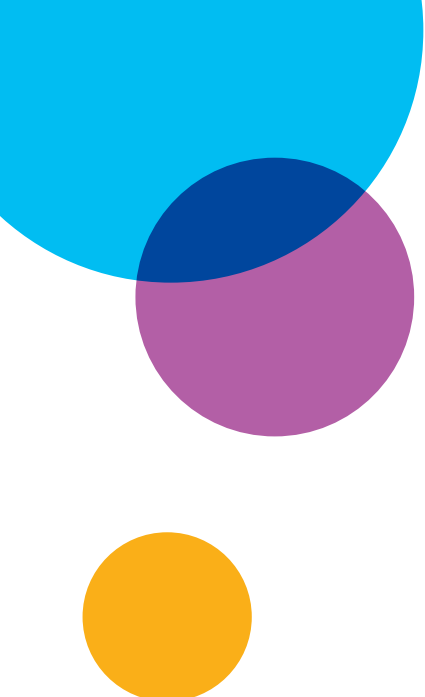
Show **slide 11** and explain that the different types of algorithms can be optimized for different effects. For example, a video site's "Up Next" or "For You" algorithm might be optimized for one of these purposes:

- Watch time: Trying to make users watch as much of each video as possible
- Engagement: getting users to comment on, Like and reply to as many videos as possible
- Stickiness: Trying to make sure that users keep watching videos instead of leaving the platform
- Virality: Encouraging users to share the videos they watch with as many people as possible, or
- Daily use: Making sure that users come back to the app often

Advance to **slide 12**. Ask students to think of an app they use frequently that uses algorithms to recommend or deliver content. Ask:

- What do you think that app is optimized for?
- Why do you think that?

After students have discussed this for a few minutes, show **slide 13** and explain that algorithms rank and weight different factors such as views, likes, freshness, shares, links and subscribers in order to decide which videos to recommend to you. Each of these categories will have different weights, depending on the platform and what it has been optimized for.



Show **slide 14** and instruct students to think again about the app they considered just before and ask:

- What inputs do you think its algorithm looks for?
- Which do you think are weighted most heavily?
- Why do you think that?

Have students discuss these questions for a few minutes.

INPUTS

Show **slide 15** and point out that we don't always know that we're giving an algorithm information. Explain that there are explicit inputs, or those inputs which we know about and have a choice about, such as the "Like" button which tells the algorithm to show you more content like this.

Advance to **slide 16** and explain that there are implicit inputs, or ones you provide without knowing you're doing it, such as how long you watch a video. Algorithms also draw on things like your browsing history to determine what you're likely to respond to, and e-commerce sites often set higher prices if you're using an Apple device.

MACHINE LEARNING ALGORITHMS, OR ARTIFICIAL INTELLIGENCE

Advance through **slides 17**. Tell students that now that you have an understanding of how algorithms work, you're going to talk about artificial intelligence. Ask students what the term "AI" means to them.

After eliciting a few responses, advance to **slide 18** and explain that what we call AIs are *machine learning algorithms*, and that they process huge amounts of data to find connections that people wouldn't necessarily notice, then use those to make inferences, or guesses. If a guess seems to work out, it will become a rule.

Show **slide 19** and ask, based on what students know about algorithms and AI, do they think:

- AIs have feelings?
- AIs can feel hurt or left out?
- AIs can choose to do or not do things?

- AIs are smart?
- AIs always give you correct information?
- AIs know if they are giving you correct information or not?
- AIs can think about whether something is right or wrong?

Explain to students that many people are unsure about the answers to these questions. In one study, a quarter of eleven-year-olds thought that “Alexa” (the name of Amazon Echo’s voice assistant, a smart speaker that uses AI) has feelings and can think for itself, and another third thought that these were “maybe” true. Share that AI tools certainly seem to be intelligent (thus the term “artificial intelligence” and may even seem to have feelings. They even seem smart, and they do make decisions – for example, they may refuse to do something you tell them to, and if you ask them a question they will “choose” what to include or leave out of the answer.

Advance to **slide 20** and explain that, while the good and the harm done by any technology is primarily the result of how it is used, the *nature* of the technology also shapes how we use it. Use algorithms as an example: first, because of how they operate and how we interact with them is often invisible to us; and second, because machine learning algorithms resist our understanding.

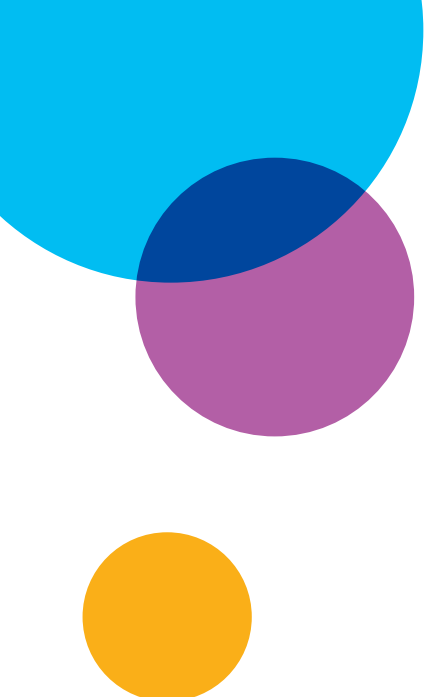
ARE ALGORITHMS ALWAYS FAIR?

Move to **slides 21** and tell students that using algorithms makes decisions more consistent and that while it *can* make decisions fairer (because you’re applying the same standards to everyone), if you’re not careful, it can also make it harder to realize when an algorithm isn’t fair.

Show slides 22-23 and explain the theory behind the 2007 KP (kidney priority) = YL (years of life) algorithm – that, in theory, the more years of life a kidney could give a person determined priority on the list. This maximized the longevity of each kidney assigned in terms of years of life.

Show **slide 24** and ask students

- Does this seem reasonable to you?
- Does it seem fair?
- Why or why not?



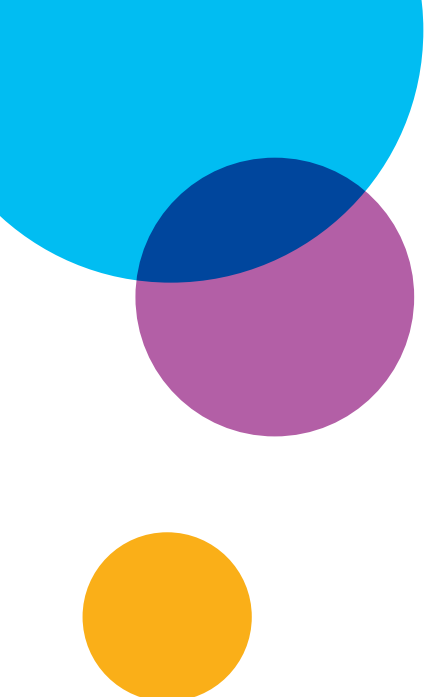
Let students discuss this for a few minutes and then advance to **slide 25**. Tell students that the older someone is, the fewer extra years of life a kidney is likely to give them. According to this algorithm, that would move them down the KP list. Ask students if it's fair to give younger people a better chance at getting a kidney. Now point out that different groups of people have different life expectancies: for example, in many countries, groups that are discriminated against or have less access to medical care, such as Black or Indigenous people, have shorter life expectancies. Ask students if it is fair for people who already have a higher life expectancy to have first priority for kidney transplants.

After students have discussed this question for a few minutes, advance to **slide 26** and explain that the people who were designing this algorithm realized that it would need to consider all those questions. For instance, the kidney transplant algorithm also considers how healthy the donor kidney is, so that younger patients are not necessarily more likely to get a kidney, but they will get a healthier kidney. Explain that, while algorithms may look objective and neutral, the fact that they're built using real-world data means that they can reinforce real-world biases.

Show **slide 27** and share that these real-world biases are expressed in algorithms. For example, AI image-generators like Midjourney and DALL-E are trained on more stock photos than on actual photos, so the images they generate reflect the conscious and unconscious choices of the stock photo companies. As a result, these algorithms are actually more biased than the real world.

Advance to **slides 28**. Explain that AI algorithms are trained (rather than programmed). This means they're given a goal (like deciding who should get a kidney), then given a data set (like a list of people who got kidney transplants in the past and how long they lived after), and then made to look for patterns in that data that they use to make a decision. Share that this is why AI is sometimes called a "black box": because unlike the kidney-transplant algorithm, we don't always know why an AI made a decision. Even the people who designed it might not know.

Show **slide 29** and tell students that this can make biases even harder to spot and have them consider the following example: one algorithm scanned thousands of successful and unsuccessful resumes and concluded that the most important qualities of a job applicant were that they played lacrosse and were named "Jared."



Advance to **slide 30** and tell students that, while the people who designed that algorithm realized it was biased and decided not to use it, if they hadn't, you might never have known that the reason you didn't get a job interview was because your name wasn't Jared and you didn't play lacrosse.

AI AND DECISION-MAKING

Show **slide 31** and reiterate that algorithms and AI are often used to make choices. Sometimes, like with kidney transplants, it's to make those decisions more consistent and more fair; other times, like when recommending videos, it's because millions of decisions have to be made every second.

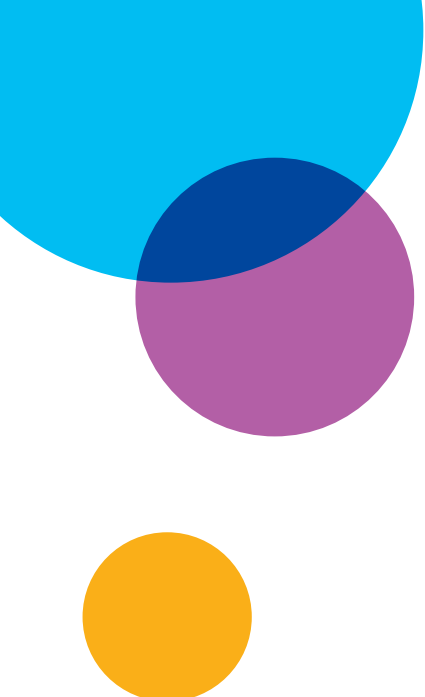
Advance to **slide 32** and point out that another thing that makes AIs different from simple algorithms is that they can change themselves. Use this example: an algorithm optimized to give you a dinner you enjoyed the most might analyze how much you ate of different things it fed you and adjust what you got based on that. Ask students how this might be problematic. What happens if you eat less broccoli? What happens if you eat more ice cream?

Show **slide 33** and explain that as time goes on, you will get more and more ice cream and less and less broccoli. Eventually, you will get only ice cream and no broccoli. If there are more than two elements of your dinner, eventually you will only get your favorite one and nothing else.

Show **slide 34** and explain that this is because algorithms are optimized to give you things that will keep you using the app or platform that uses them. You might like getting less broccoli and more ice cream, but it's not necessarily good for you.

Now, advance to **slide 35** and explain that this process is called a *feedback loop* and that students might see this in action on platforms like YouTube and Netflix, where the algorithm has gotten so precise that it recommends more and more of the same stuff. Content creators, from social media influencers to steaming production companies, create content they think the algorithm will push, which makes all their content more and more similar as the algorithm zeroes in on what most users will respond to.

Show **slide 36** and ask students for some examples of "ice cream," or things that aren't good for them in large amounts (or at all), that their algorithm shows them.



After students have given some examples, share some more: clickbait, fake news, things that make them upset, things that make them feel bad about themselves, or things that make them keep watching when they know they shouldn't.

Show **slide 37** and ask students for examples of things they see in their feed that make them feel better.

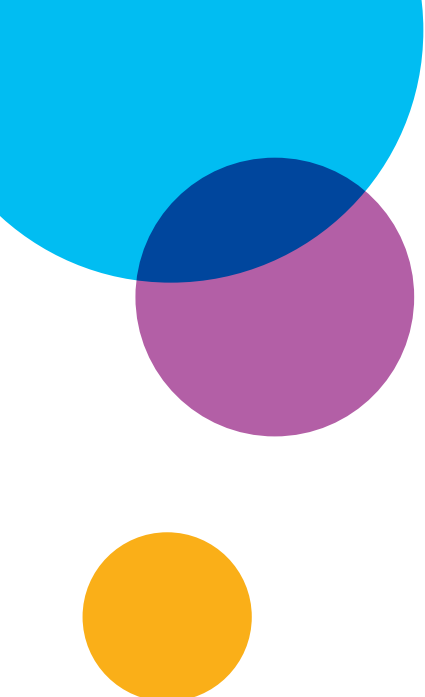
After students have given some examples, share some more: real news, reliable information, things that make them feel good about themselves, or things that encourage them to be healthy.

Advance to **slide 38** and explain that they are going to consider how algorithms get input, or the data they need to make decisions.

Show **slide 39** and tell students to take auto insurance for example: the more likely you are to get into an accident, the more you have to pay for insurance. Explain to students that insurance companies already set prices based on demographic data: men pay more than women for car insurance, on average, because they have more accidents. Now, have students imagine that they are designing an algorithm for an insurance company that wanted to know who was more likely to get into an accident. What if your AI knew what videos people had watched on TikTok or YouTube, what they had bought at e-commerce sites like Amazon, or what they had searched for on Google, and could match that with whether or not people had car accidents?

Show **slide 40** and explain how AI might use this information – it might match search, viewing or shopping history with more or less risky behaviors, like buying a skateboard or watching skateboarding videos. Describe how this might positively impact those who are seen as lower risk, but negatively impact those who were seen as being higher risk.

Advance to **slide 41** and tell students that the data might let the AI make more accurate decisions than ones based on very broad categories like gender, but the connections it makes might be inaccurate if the data didn't mean what it seemed to mean (for instance, if someone bought a skateboard as a gift) or if the connection between one kind of risk behavior and another wasn't accurate (for instance, if you like risky sports but are a careful driver).



Show **slide 42**. Tell students that this is how AIs work. They either use data they collected about you (like prompts entered into a chat AI or information your computer automatically sends like your location and IP address), information the company bought (either from another company or from data brokers that buy data from lots of places and put it together into profiles), or information shared between different parts of the same company (like how Google can use your YouTube views to decide which search results to show you).

Show **slide 43** and summarize how algorithms work: they make decisions based on what they know about you, but also on what they *think* they know about you.

GENERATIVE AI

Advance to **slide 44** and tell students that you are now going to learn about generative AI. Generative AI, like Dall-E, ChatGPT, and Midjourney, work the same way as other machine learning algorithms, but they draw on much more data and have much more complex models than even the AIs used by TikTok, YouTube, or Google.

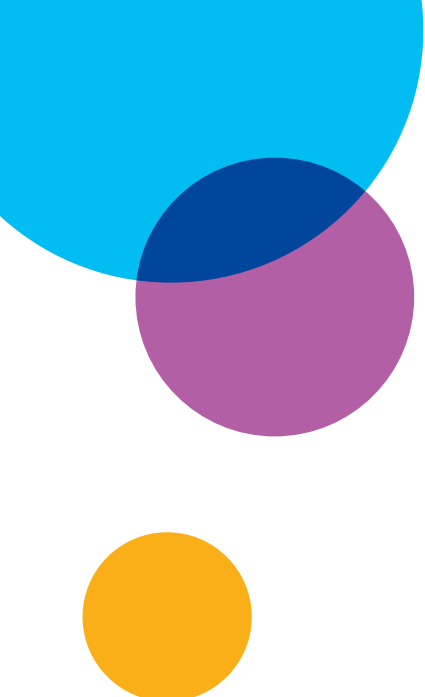
Show **slide 45** and tell students that you will be using ChatGPT – a large language model – as your example.

Show **slide 46**. Remind students that a model is a machine learning algorithm that, rather than being programmed, is trained on a large amount of writing. They find patterns in these to create a model of how language works.

Advance through **slides 47-48** while you explain that ChatGPT can read and write fluently at the level of sentences, paragraphs and even full articles. It does this mostly by looking at how similar or different words are in different ways or “dimensions.”

For example, if you were to consider just two dimensions, roundness and redness, an apple and a fire truck might be far apart on roundness but close together on redness, while a baseball would be close to the apple in terms of roundness but far away in terms of redness.

Chat AIs make guesses by “looking” along different dimensions: if it started at “king” and looking further along the way the “female” dimension it would see “queen,” while if it looked down the youth dimension it might see “prince,” and looking in both directions might



lead to “princess.” Tell students that this lets the AI make better guesses about what words should follow each other based on other parts of the sentence or paragraph.

Advance through **slides 49-53** to guide students through the following example:

- If you were to write “Frida had a drink of chocolate” ...
- Autocomplete might always suggest that the next word after “chocolate” should be “chips” because that’s what follows it most often in the training set.
- On the other hand, if you ask ChatGPT “What kind of chocolate did Frida drink?” ...
- It might spot the word “drink” and then look from “chocolate” along the liquid dimension and find that the nearest word in that direction was “chocolate milk.”

Show **slide 54**. Extend the example to other dimensions large language models can consider – suppose you’re starting with chocolate and looking along the “brown” and “liquid” dimensions. If you add a third dimension – let’s say “edible” – you’re already reaching the point where it’s hard to represent the relationship graphically, and the limits of what most people can hold in our minds.

Advance to **slide 55**. Explain to students that this is where “large” comes in: in ChatGPT, each word is given a value in up to 96 dimensions, and it does more than 9000 operations every time it guesses a new word. It was trained on a data set of around 500 billion words. That kind of scale is how chat AIs are able to mimic real language and conversations.

Advance to **slide 56** and share that this is also how AIs like DALL-E are able to make convincing images. Highlight the importance of remembering that generative AI still works by making guesses, based on probability and on similarity. The size of the training set and the sophistication of the models means they can make fluent content, and if the training set included the right answer to your question, you might get accurate information – but large language models are also prone to *hallucinations*, where they will tell you with perfect confidence about things that don’t exist.

FOUR PRINCIPLES OF ALGORITHMS AND AI

Show **slide 57** and review the four principles that you covered: algorithms can reinforce real-world biases, we don't always know how AIs make decisions, AIs give you what they think you want (not what's good for you), and AI makes decisions based on what it thinks it knows about you.

Show **slide 58** and ask students:

- What kind of biases might be found in the texts AIs were trained on? (Think about social media posts, or in old books.) How might that affect its responses?
- Should people who make AIs have to be able to explain how they work? How does their “black box” quality make it hard for designers to make safeguards or “guardrails” that might prevent some of the other problems?
- Are there topics or information you might want a chat AI to tell you, but that wouldn't be good for you? What could go wrong if you asked a chat AI for advice, or told them about a personal problem? What could go wrong if you used them as a source of information?
- How do you think AIs get information about you? Do you think that's fair? Do you think it's accurate?
- What kinds of guardrails should be required for AI developers? In which decisions should we need a human in the loop?

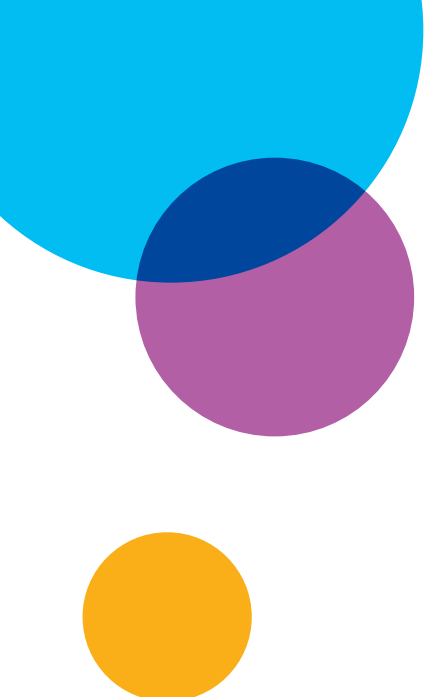
ACTIVITY: ALGORITHM AIS AND BIAS

After students have discussed these questions, distribute the assignment sheet *Algorithm AIs and Bias* and tell students that they will be creating an infographic on a field that uses algorithms to make decisions, much like the kidney transplant example from this lesson.

The infographic will explore how these AIs work and provide examples of how biases manifest in their specific AI system. Depending on technology available, you may permit students to create their infographic on programs like Canva, Google Drawings, PowerPoint or another [alternative](#).

In small groups (of 2-3), have students select one of the following fields that uses AI algorithms:

- Recommendation algorithms on video sites or social networks (YouTube Up Next bar, TikTok For You page, etc.)

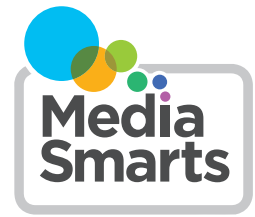
- 
- Sorting algorithms on search engines
 - AI algorithms in retail
 - Medical AI
 - Facial recognition
 - Hiring tools
 - Criminal justice algorithms

Tell students that the infographic should include a definition and explanation of AI algorithms, visual representations of how bias can occur in AI algorithms, real-world examples and their impact, and a reflection on how this impacts their own lives.

Depending on the time available, you may choose to host a gallery walk, where students can engage with their peers' projects. This may take place in class, or virtually through a Learning Management System or a program like VoiceThreads. Students may choose to include images in their infographic. To that end, the assignment sheet has some suggestions for sources of copyright-free images for students to use, if desired.

When students have completed their infographic (or after the gallery walk), they then write a paragraph responding to two or more of the following prompts:

- Did this presentation share something that surprised you or that you didn't expect?
- How does the way this AI algorithm works impact society as a whole?
- Are you impacted personally by the bias inherent in this AI? How does that make you feel? If you are not directly impacted, might you view it differently than someone who is affected?
- How might the bias presented here impact your interaction with this type of product?



UNPACKING THE BLACK BOX: EXPLAINING ALGORITHMS AND AI

Algorithm AIs and Bias Assignment Sheet

.....

For this project, you will create an infographic on a field that uses algorithms to make decisions. Much like the kidney transplant example from our lesson, you'll investigate how algorithms work, their impact, and any biases that may manifest.

As a group, choose one of the following fields that use AI algorithms:

- Recommendation algorithms on video sites or social networks (YouTube Up Next bar, TikTok For You page, etc.)
- Sorting algorithms on search engines
- AI algorithms in retail
- Medical AI
- Facial recognition
- Hiring tools
- Criminal justice algorithms

Research how algorithms are used in your chosen field. Consider how the algorithms work (what data do they use and what decisions do they help make?), real-world examples (what specific instances have these algorithms been used?), and biases and impacts (what known biases exist and how do they impact others?).

When planning your infographic, ensure that you include:

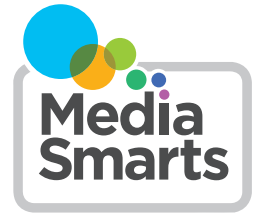
- A clear definition of AI algorithms as it relates to your field and an explanation of how they function.
- Examples of the algorithm in use, where biases have impacted decisions or outcomes, and the broader implications of these examples.
- Reflect on how bias in AI algorithms might impact your own life or the lives of people you know. (Consider the decision-making processes that influence your daily life!)

Use visuals to make the infographic engaging and easy to understand! Your poster may contain images, if you so choose. Remember, you must give credit and ensure that you have a legal right or license to use photos that you did not create! Here are some copyright-cleared or Creative Commons image sources for your use:

- [pexels.com](https://www.pexels.com)
- archive.org
- pics4learning.com
- openclipart.com
- rawpixel.com/public-domain

When you have completed your infographic, write a paragraph responding to two or more of the following prompts:

- Did this presentation share something that surprised you or that you didn't expect?
- How does the way this AI algorithm works impact society as a whole?
- Are you impacted personally by the bias inherent in this AI? How does that make you feel? If you are not directly impacted, might you view it differently than someone who is affected?
- How might the bias presented here impact your interaction with this type of product?



UNPACKING THE BLACK BOX: EXPLAINING ALGORITHMS AND AI

Algorithm AIs and Bias Planning Sheet

.....

MY GROUP'S FIELD THAT USES AI ALGORITHMS:		
How the algorithms work What data do they use? What decisions do they help make?	Real-world examples: Find specific instances where these algorithms have been used	Biases: Investigate any known biases in these algorithms and their effects on people or outcomes